Restricted Bayesian Optimal Design

By

Han Son Seo and Kinley Larntz
School of Statistics, University of Minnesota

Technical Report No. 574

February, 1992

Restricted Bayesian optimal design

By HAN SON SEO and KINLEY LARNTZ

School of Statistics, University of Minnesota, St. Paul, MN 55108, U.S.A.

SUMMARY

We consider the experimental design problem of selecting values of design variables x for observation of a response y that depends on x and on model parameters θ. The form of the dependence may be quite general, including all linear and nonlinear modeling situations. The goal of the design selection is to efficiently estimate functions of θ. Three new criteria for selecting design points x are presented. The criteria generalize the usual Bayesian optimal design criteria to situations in which the prior distribution for θ may be uncertain. We assume that there are several possible prior distributions, one of which may be considered as more plausible than the others. Designs that minimize the new criteria have the characteristic of being robust with respect to choice of prior distribution. The new criteria are applied to the nonlinear problem of designing to estimate the turning point of a quadratic equation. We give both analytic and computational results illustrating the robustness of the optimal designs based on the new criteria.

Some key words: A-optimality; Nelder-Mead algorithm; robust design; turning point problem.

1. INTRODUCTION

In optimal design problems, issues of robustness can be addressed for two specific questions: model robustness and parameter robustness. Model robustness concerns whether the model form assumed in the design process is correct. Is the error distribution normal or t? Is the regression linear or quadratic? Parameter robustness concerns the sensitivity of nonlinear designs to the assumed values of the model parameters when designs are based on the investigator's best guess for model parameters. In this paper we deal with both robustness aspects by specifying our design assumptions in several prior distributions, one of which is assumed to be more plausible than the others. By choice of prior, aspects of both model robustness and parameter robustness may be addressed. Stigler(1971) treated an optimal design problem in which he designed for a polynomial of a given degree, but required that the design be robust to the degree of the polynomial. He introduced new D- and G-optimality criteria which may be considered as compromises between the incompatible goals of efficient inference about the regression function under a particular model and checking the adequacy of that model. Studden(1982) and Lau(1988) used a technique involving canonical moments to find robust D-optimal designs for the same problem as Stigler. Cook and Nachtsheim(1982) developed a new design criterion that generalizes linear optimality to a situation in which, a priori, the exact form of the regression model need not to be known.

In designing an optimal experiment to estimate a nonlinear combination of coefficients of a linear model, or, in general, in designing experiments for nonlinear models, the efficiency of a design depends on the values of the unknown parameters. A common approach for handling this difficulty is to design the experiment to be optimal for a best guess of the parameter values. Chernoff (1953) termed this approach "locally optimal" design. A natural generalization is to use a prior distribution on the unknown parameters rather than a single guess. But it may happen that there are several plausible prior distributions. DasGupta and Studden in a Purdue University technical report considered several optimal robust criteria in normal linear models. They formulated uncertainty of the prior in terms of having a family of priors in place of a single prior. They used a family of conjugate priors and a family of priors induced by a metric on the space of nonnegative measures.

Motivated by the model robust criteria in the papers mentioned above, we present new criteria which quantify robustness in the case of several plausible prior distributions. In this paper we assume one prior is more plausible than the others. We suggest three new criteria for this case. We apply the criteria to the problem of estimating the turning point of a quadratic regression. Section 2 briefly reviews Bayesian optimal design theory. Section 3 gives three new criteria. Section 4 illustrates application of the criteria to the nonlinear problem of estimating the turning point of a quadratic regression.

2. BAYESIAN OPTIMAL DESIGN THEORY

Consider the following experimental design problem :

Suppose y is a random variable with density function $p(y|\theta,x)$ depending on parameters $\theta^T = (\theta_1,...,\theta_p)$ and design variable x. The design variable is restricted to an experimental region X. The experimental design problem is to choose N values of $x \in X$ with the goal of estimating functions of θ. Denote the observed variables as $y^T=(y_1,...,y_N)$ and the design variables as $x^T=(x_1,...,x_N)$. Denote the prior distribution for θ as $\Delta(\theta)$ with density $\delta(\theta)$.

The appropriate Bayesian analysis for estimating θ constructs the exact posterior distribution which has density proportional to $p(y|\theta,x)\delta(\theta)$. For most realistic models computation of the exact posterior distribution is intractable and asymptotic approximations are used. Under easily satisfied assumptions the posterior distribution of θ is approximately a multivariate normal distribution (Berger, 1985, p. 224) with mean the maximum likelihood estimate, $\hat{\theta}$. The variance-covariance matrix is the inverse of the observed Fisher information matrix, i.e.,

$$V(\theta) = I(\hat{\theta},\eta)^{-1}$$

$$\text{where} \quad [I(\hat{\theta},\eta)]_{ij} = -\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(p(y|\theta,x))\right)_{\theta=\hat{\theta}}$$

$$= -\sum_{i=1}^{N}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(p(y|\theta,x_i))\right)_{\theta=\hat{\theta}}$$

For convenience, we expand the definition of a design to include any probability measure η on X. We define the normalized information matrix $I(\hat{\theta},\eta)$ by

$$[I(\hat{\theta},\eta)]_{ij} = -\int\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log(p(y|\theta,x))\right)_{\theta=\hat{\theta}}\eta(dx)$$

For any function of the parameters, $g(\theta)$, define the loss function for estimating $g(\theta)$ with $\tilde{g}$ as $L(\tilde{g}, g(\theta))$. Posterior expected loss can be approximated using the above approximate distribution of θ. Here, we take as our loss function the usual squared error loss $L(\tilde{g}, g(\theta))=(\tilde{g}-g(\theta))^2$. In this case, the usual criterion for choosing an optimal design corresponds to the approximate expected posterior variance of $g(\theta)$. If several functions of θ are of interest, Chaloner and Larntz (1989) suggest a more general form of the criterion using the expected weighted trace of the product of a symmetric matrix and the inverse of the information matrix. The criterion is

$$\phi(\eta) = E_\theta(\mathrm{tr}B(\theta)I(\theta,\eta)^{-1})$$

where $B(\theta) = B^{\frac{1}{2}}(\theta)\, B^{\frac{1}{2}}(\theta)^T$ is a symmetric p by p matrix and I is the Fisher information matrix. For a measure η for which $I(\theta,\eta)$ is singular for a θ value with non-zero prior probability we define $\phi(\eta)$ to be ∞. Specifically if interest centers on k functions, $g_1(\theta),\ldots,g_k(\theta)$, then

$$
B^{\frac{1}{2}}(\theta) = \begin{pmatrix} \frac{\partial g_1(\theta)}{\partial \theta_1} & \frac{\partial g_2(\theta)}{\partial \theta_1} & \cdots & \frac{\partial g_k(\theta)}{\partial \theta_1} \\ \frac{\partial g_1(\theta)}{\partial \theta_2} & \frac{\partial g_2(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_k(\theta)}{\partial \theta_2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial g_1(\theta)}{\partial \theta_p} & \frac{\partial g_2(\theta)}{\partial \theta_p} & \cdots & \frac{\partial g_k(\theta)}{\partial \theta_p} \end{pmatrix}
$$

If linear combinations of the θ_i are of interest then $B(\theta)$ does not depend on θ and is a matrix of fixed values. If non-linear combinations of the θ_i's are of interest then $B(\theta)$ has entries which are functions of θ. An optimal design is a design measure η which minimizes $\phi(\eta)$.

3. ROBUST BAYESIAN DESIGN CRITERIA

Consider the general Bayesian design problem with Δ, the prior distribution for the parameter θ. We now want to consider having several (n, say) possible prior distributions for θ. Denote these by Δ_i, $i=1,\ldots,n$. For each prior distribution, let $\Phi(\eta,\Delta_i)$ denote a certain function of interest for design η evaluated for prior Δ_i. The goal is to minimize $\Phi(\eta,\Delta_i)$ over η.

Definition 3.1

We call η^* a B-optimal design for the prior Δ if η^* minimizes $\Phi(\eta,\Delta)$ among all designs η.

To assess the relative worth of a design η for prior Δ, we use the efficiency.

Definition 3.2

The efficiency of a design η with respect to prior Δ is defined by

$$\mathrm{Eff}(\eta,\Delta) = \frac{\min_{\chi} \Phi(\chi,\Delta)}{\Phi(\eta,\Delta)} = \frac{\Phi(\eta^*,\Delta)}{\Phi(\eta,\Delta)}$$

where η^* is a B-optimal design for prior Δ.

Assume that among several possible priors one is considered more plausible than the others. If we denote the more plausible prior as Δ_1 and the others as $\Delta_{21}, \ldots, \Delta_{2n}$, we may want to find a design that minimizes $\Phi(\eta,\Delta_1)$ subject to $\Phi(\eta,\Delta_{21}), \ldots, \Phi(\eta,\Delta_{2n})$ being not too big. We call Δ_1 the "major prior" and $\Delta_2=\{\Delta_{21}, \ldots, \Delta_{2n}\}$ the "set of minor priors."

For the above situation, we define three criteria for robust Bayesian optimal designs in the following manner.

Definition 3.3

η_0 is a k-restricted B_1-optimal design for major prior Δ_1 against the set of minor priors $\Delta_2=\{\Delta_{21}, \ldots, \Delta_{2n}\}$ if η_0 minimizes $\Phi(\eta,\Delta_1)$ among all designs η satisfying

$$\max_i \Phi(\eta,\Delta_{2i}) \le k\Phi(\eta,\Delta_1) \qquad \text{for given } k \ge 1.$$

B_1-optimal design restricts the choice of optimal design for the major prior to those designs for which each minor prior has criterion value less than k times the criterion value for the major prior. DasGupta and Studden in their Purdue University technical report mentioned this criterion. Also Stigler(1971) used a frequentist version of this criterion to construct designs robust to the degree of a polynomial regression.

If $\Phi(\eta,\Delta_{2i})$'s are generally small compared to $\Phi(\eta,\Delta_1)$, then for a small $k > 1$, many designs satisfy the condition. So the B_1-optimal design may not differ much from the B-optimal design and hence may not be robust. Note that for k large enough, the B_1-optimal design will be the same as the B-optimal design since the criterion becomes increasingly less restrictive as k increases. For this reason we may want to define robust optimal design in terms of efficiency directly.

Definition 3.4

$\tilde{\eta}$ is a t-restricted B_2-optimal design for major prior Δ_1 against the set of minor priors $\Delta_2=\{\Delta_{21}, ..., \Delta_{2n}\}$ if $\tilde{\eta}$ minimizes $\Phi(\eta,\Delta_1)$ among all designs η satisfying

$$\min_i \mathrm{Eff}(\eta,\Delta_{2i}) \geq t\ \mathrm{Eff}(\eta,\Delta_1) \text{ for given } t \in (0,1).$$

B_2-optimal design restricts the choice of optimal design for the major prior to those designs having efficiency for each minor prior at least t times the efficiency of the design for the major prior. An alternative criterion put restrictions on the absolute efficiency.

Definition 3.5

$\hat{\eta}$ is an s-restricted B_3-optimal design for major prior Δ_1 against the set of minor priors $\Delta_2=\{\Delta_{21}, ..., \Delta_{2n}\}$ if $\hat{\eta}$ minimizes $\Phi(\eta,\Delta_1)$ among all designs η satisfying

$$\min_i \mathrm{Eff}(\eta,\Delta_{2i}) \geq s \text{ for given } s \in (0,1).$$

B_3-optimality requires that the efficiency of the design be at least s for all minor priors.

4. EXAMPLE: TURNING POINT PROBLEM

4.1 Introduction

This section presents results on finding various B-optimal designs for estimating the turning point of a quadratic regression. The model under consideration is quadratic regression where observations y_i are taken at design points x_i. The observations are such that

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

where $\beta = (\beta_0, \beta_1, \beta_2)$ are unknown coefficients and the chance errors e_i are assumed independent, normally distributed with mean zero and variance σ^2. The purpose of the experiment is to estimate the turning point

$$\gamma = -\frac{\beta_1}{2\beta_2},$$

the value of x at which the expected value of y is a maximum if β_2 is negative, or is a minimum if β_2 is positive. Consider collecting observations $y_1, ..., y_N$ corresponding to N chosen levels $x_1 \le x_2 \le ... \le x_N$ of x. Suppose the x's must be in the finite interval [a, b]. Without loss of generality we may restrict the x_i to be in the interval [-1, 1]. Given N, our problem is to select the design $\mathbf{x} = (x_1, ..., x_N)$, $x_i \in [-1, 1]$ for i = 1 ... N. The B-optimal design depends on β_0, β_2 and γ only through the first two moments of distribution of γ. For convenience, we summarize the prior for γ as the vector, $\Delta = (E(\gamma), Var(\gamma))$. Suppose $E(\gamma)=m$ and $var(\gamma)=v$ which are known. That is,

$$\Delta = (m, v).$$

For large sample size, the expected posterior variance of γ is proportional to

$$\Phi(\eta,\Delta) = E_\gamma(\text{tr } B(\gamma)\, I(\theta,\eta)^{-1})$$

$$\text{where } B(\gamma) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2\gamma \\ 0 & 2\gamma & 4\gamma^2 \end{pmatrix}$$

From Mandal (1978),

$$\Phi(\eta,\Delta) = \frac{1}{d^2}\left[\frac{1}{\mu_2} + \frac{\frac{4}{d^2}v + \left\{\frac{2}{d}(m-c) - \frac{1}{\mu_2}(\mu_3' - \mu_2'\mu_1')\right\}^2}{\mu_4' - \mu_2'^2 - \frac{1}{\mu_2}(\mu_3' - \mu_2'\mu_1')^2}\right]$$

where

$$z_i = \frac{2x_i - x_1 - x_N}{x_N - x_1}, \quad \mu_r' = \frac{1}{N}\sum_{i=1}^{N} z_i^r, \quad \mu_r = \frac{1}{N}\sum_{i=1}^{N}(z_i - \mu_1')^r, \quad c = \frac{x_1 + x_N}{2}, \quad d = \frac{x_N - x_1}{2}.$$

For the special case of the prior centered at 0, i.e. m=0, the B-optimal design is given by Mandal (1978).

Theorem 4.1 (Mandal)

A B-optimal design for $\Delta = (0,\ \nu)$ is

$$x_i = \begin{cases} -1 & \text{with proportion} \quad \dfrac{\mu_2'^*}{2} \\ 0 & \text{with proportion} \quad 1-\mu_2'^* \\ 1 & \text{with proportion} \quad \dfrac{\mu_2'^*}{2} \end{cases}$$

where $\mu_2'^* = \{1+2(\nu^{-1}+4)^{-\frac{1}{2}}\}^{-1}$

For further details on the B-optimal design, see Chaloner(1989).

Using theorem 4.1, we find B_1-, B_2- and B_3-optimal designs analytically for certain restriction values k, s and t for the turning point problem with the major prior having mean at the center of the design region.

4.2 B_1-optimal design

First we note that if a restricted value k is large enough, the B_1-optimal design coincides with the B-optimal design. Specifically if $k \geq k^*$,

$$k^* = \max_i \frac{\Phi(\eta_1^*,\Delta_{2i})}{\Phi(\eta_1^*,\Delta_1)}$$

where η_1^* is the B-optimal design for prior distribution Δ_1, then the k-restricted B_1-optimal design is the B-optimal design. For smaller values of k, we can determine the B_1-optimal design analytically if the major prior has mean 0 and $k > \bar{k}$,

$$\tilde{k} = \frac{1 + 4 \max_i (\nu_{2i} + m_{2i}^2)}{1 + 4\nu_1} .$$

Specifically, for $k \geq 1$ and $\tilde{k} < k < k^*$, a k-restricted B_1-optimal design is

$$x_i = \begin{cases} -1 & \text{with proportion} \quad \dfrac{\tilde{\mu}_2'}{2} \\ 0 & \text{with proportion} \quad 1 - \tilde{\mu}_2' \\ 1 & \text{with proportion} \quad \dfrac{\tilde{\mu}_2'}{2} \end{cases}$$

where $\tilde{\mu}_2' = 1 - 4\{k\nu_1 - \max_i (\nu_{2i} + m_{2i}^2)\}/(1-k)$.

Note that the design has only 3 support points and is symmetric.

Example: To illustrate the robustness of B_1-optimal design, consider the following example. We have four priors, each with mean 0, but with variances 0.03, 0.07, 0.15 and 0.9, respectively. Figure 1 graphs the efficiencies of various B_1-optimal designs and compares them to the B-optimal design for major priors with variances 0.03, 0.07 and 0.15. Looking at Figure 1(a), note that the B-optimal design for ν_1=0.03 has efficiency less than 80% when the true variance is 0.9. In contrast the B_1-optimal design with k=8.50 has efficiency 93.5% for variance 0.9 while maintaining efficiency 96.0% when ν_1=0.03. The efficiencies for variances 0.07 and 0.15 are virtually 100%. For major prior ν_1=0.9, the B_1-optimal design is same as the B-optimal design for all values of k. Note that the gain from using the B_1-optimal design is smaller when the major prior is ν_1=0.07 and ν_1=0.15. Nonetheless the efficiencies are less extreme across the range of prior distributions than the unrestricted B-optimal designs.

4.3 B_2-optimal design

When t is small, the B_2-optimal design is the same as the B-optimal design. Specifically if $0 < t \leq t^*$,

$$t^* = \min_i \mathrm{Eff}(\eta_1^*, \Delta_{2i})$$

where η_j^* are B-optimal designs for prior distribution Δ_j then the B-optimal design is also the t-restricted B_2-optimal design. When the major prior has mean 0 a B_2-optimal design can be found analytically for t-values greater than t^* and satisfying a certain condition. We first give some notation. For a fixed value of t, B_i is defined as

$$B_i = t^{-1} \frac{\Phi(\eta_{2i}^*, \Delta_{2i})}{\Phi(\eta_1^*, \Delta_1)} \ .$$

And for a minor prior Δ_{2i}, $f_i(d)$ is a function of d defined as

$$f_i(d) = \frac{1}{d^2} \frac{4\,(v_{2i} + m_{2i}^2 - B_i v_1)}{(1-B_i)} + 1\,.$$

For a fixed value of t, we define sets of minor priors as following :

$\Omega = \{\Delta_{21}, \ldots, \Delta_{2n}\}$

$D(t) = \{\Delta_{2i} \mid v_{2i} + m_{2i}^2 > v_1\}$, $E(t) = \{\Delta_{2i} \mid v_{2i} + m_{2i}^2 \le v_1\}$

So $D(t) \cup E(t) = \Omega$.

$D_1(t) = \{\Delta_{2i} \mid\ t \in (t^*, t_{2i}^*) \text{ and } \Delta_{2i} \in D(t)\ \}$

$D_2(t) = \{\Delta_{2i} \mid\ t \in (t_{2i}^*, t_{3i}^*) \text{ and } \Delta_{2i} \in D(t)\ \}$

$E_1(t) = \{\Delta_{2i} \mid\ t \in (t^*, t_{1i}^*) \text{ and } \Delta_{2i} \in E(t)\ \}$

$E_2(t) = \{\Delta_{2i} \mid\ t \in (t_{3i}^*, t_{2i}^*) \text{ and } \Delta_{2i} \in E(t)\ \}$

$E_3(t) = \{\Delta_{2i} \mid\ t \in (t_{1i}^*, t_{3i}^*) \text{ and } \Delta_{2i} \in E(t)\ \}$

where $t_{1i}^* = \dfrac{\Phi(\eta_{2i}^*, \Delta_{2i})}{\Phi(\eta_1^*, \Delta_1)}$, $t_{2i}^* = \dfrac{v_1\, t_{1i}^*}{v_{2i} + m_{2i}^2}$, $t_{3i}^* = \dfrac{(4v_1+1)\, t_{1i}^*}{4\,(v_{2i} + m_{2i}^2) + 1}$.

We also define f_D and f_E as

$$f_D = \min_{i \in I(D_2)} f_i(1), \quad f_E = \max_{i \in I(E_2)} f_i(1)$$

where $I(D_2) = \{i \mid \Delta_{2i} \in D_2(t)\}$ and $I(E_2) = \{i \mid \Delta_{2i} \in E_2(t)\}$.

Using the above notation we consider following cases:

(I) $D_1(t) \cup E_1(t) \cup D_2(t) \cup E_3(t) = \Omega$ and $D_2(t) \neq \phi$

(II) $D_1(t) \cup E_1(t) \cup D_2(t) \cup E_2(t) \cup E_3(t) = \Omega$, $D_2(t) \neq \phi$ and $E_2(t) \neq \phi$,

$f_D > f_E$, and $\mu_2'^* > f_D$

(III) $D_1(t) \cup E_1(t) \cup E_2(t) \cup E_3(t) = \Omega$ and $E_2(t) \neq \phi$,

(IV) $D_1(t) \cup E_1(t) \cup D_2(t) \cup E_2(t) \cup E_3(t) = \Omega$, $D_2(t) \neq \phi$ and $E_2(t) \neq \phi$,

$f_D > f_E$, and $\mu_2'^* < f_E$

where $\mu_2'^* = \{1+2(\nu_1^{-1}+4)^{-\frac{1}{2}}\}^{-1}$.

Then for case I and case II, a B_2-optimal design is

$$x_i = \begin{cases} -1 & \text{with proportion} \quad \frac{f_D}{2} \\ 0 & \text{with proportion} \quad 1 - f_D \\ 1 & \text{with proportion} \quad \frac{f_D}{2} . \end{cases}$$

For case III and case IV, a B_2-optimal design is

$$x_i = \begin{cases} -1 & \text{with proportion} \quad \frac{f_E}{2} \\ 0 & \text{with proportion} \quad 1 - f_E \\ 1 & \text{with proportion} \quad \frac{f_E}{2} . \end{cases}$$

Again the B_2-optimal design is a three point design. Figure 2 presents efficiencies for B_2-optimal design in the same manner as Figure 1 does for B_1-optimal designs. Note that for B_2-optimality, analytic results are available when the major prior has ν_1=0.9. Again note the robustness of the final design when ν_1=0.03. Similar compromise robust designs are also available when ν_1=0.9. The robustness gains when ν_1=0.07 or ν_1=0.15 are smaller than the gains when the major prior is more extreme, but, as before, the efficiencies are less extreme for the robust designs.

4.4 B_3-optimal design

If the restriction s is small enough, $s \leq s^*$,

$$s^* = \min_i \left\{ \frac{\Phi(\eta_{2i}^*, \Delta_{2i})}{\Phi(\eta_1^*, \Delta_{2i})} \right\}$$

where η_{2i}^* and η_1^* are B-optimal designs for prior distribution Δ_{2i} and Δ_1 respectively, then the s-restricted B_3-optimal design is identical to the B-optimal design. Also, for a range of s values, analytic results for B_3-optimal design may be found if the major prior mean is 0. Denote

$$A_i = \Phi(\eta_{2i}^*, \Delta_{2i})\, s^{-1}, \quad i=1,..,n.$$

Then if $s^* < s < \min\{\tilde{s}, 1\}$ where

$$\tilde{s} = \min_i \left[\Phi(\eta_{2i}^*, \Delta_{2i}) \left\{ 1+8(v_{2i} + m_{2i}^2) - \{ (1 + 8(v_{2i} + m_{2i}^2))^2 - 1 \}^{\frac{1}{2}} \right\} \right]$$

and

$$\max_i \left[\frac{1 - \{(A_i - 1)^2 - 16 A_i (v_{2i} + m_{2i}^2)\}^{\frac{1}{2}}}{4A_i} \right] \leq \min_i \left[\frac{1 + \{(A_i - 1)^2 - 16 A_i (v_{2i} + m_{2i}^2)\}^{\frac{1}{2}}}{4A_i} \right]$$

the B_3-optimal design is given by one of two cases.

Case 1:

If $\{1+2(v_1^{-1}+4)^{-\frac{1}{2}}\}^{-1} > 2^{-1} + \min_i \left[(2A_i)^{-1} [1+\{(A_i - 1)^2 - 16 A_i (v_{2i} + m_{2i}^2)\}^{\frac{1}{2}}]\right]$

then an s-restricted B_3-optimal design is

$$x_i = \begin{cases} -1 \text{ with proportion } & 4^{-1} + \min_i \left[(4A_i)^{-1} [1+\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \\ 0 \text{ with proportion } & 2^{-1} - \min_i \left[(2A_i)^{-1} [1+\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \\ 1 \text{ with proportion } & 4^{-1} + \min_i \left[(4A_i)^{-1} [1+\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \end{cases}$$

Case 2:

If $\{1+2(v_1^{-1}+4)^{-\frac{1}{2}}\}^{-1} < 2^{-1} + \min_i \left[(2A_i)^{-1} [1+\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}]\right]$

then an s-restricted B_3-optimal design is

$$x_i = \begin{cases} -1 \text{ with proportion } & 4^{-1} + \max_i \left[(4A_i)^{-1} [1-\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \\ 0 \text{ with proportion } & 2^{-1} - \max_i \left[(2A_i)^{-1} [1-\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \\ 1 \text{ with proportion } & 4^{-1} + \max_i \left[(4A_i)^{-1} [1-\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}] \right] \end{cases}$$

Note that if $s > s^*$ then

$\{1+2(v_1^{-1}+4)^{-\frac{1}{2}}\}^{-1} \neq 2^{-1} + \min_i \left[(2A_i)^{-1} [1+\{(A_i-1)^2 - 16A_i(v_{2i}+m_{2i}^2)\}^{\frac{1}{2}}]\right]$.

Figure 3 illustrates the robustness of the B_3-optimal design for the same cases we used to illustrate B_1- and B_2-optimality. For this criterion gains, in the sense of robustness of design, are substantial for all four major priors.

4.5 Numerical Results for Other Priors

When the mean of the major prior is not at the center of the design region, analytic results are intractable and finding an optimal design becomes a problem in numerical optimization. Chaloner and Larntz(1989) have found the Nelder and Mead(1965) version of the simplex algorithm effective for these problems. This is an unconstrained optimization method and we transform our problem to satisfy this. For simplicity, we assume that there are only two possible priors. With two priors, we consider four cases ;

I) Both major prior (or prior 1) and minor prior (or prior 2) are certain, i.e., have small variances.

II) Major prior is certain (small variance) but minor prior is not (large variance).

III) Major prior is not certain but minor prior is certain.

IV) Neither major prior nor minor prior is certain.

Corresponding to the above four cases, we take priors as follows in our examples.

a) $\Delta_1 = (-0.2, 0.07)$ $\quad\quad$ $\Delta_2 = (\ 0.5, 0.07)$

b) $\Delta_1 = (-0.2, 0.07)$ $\quad\quad$ $\Delta_2 = (\ 0.5, 0.3)$

c) $\Delta_1 = (-0.2, 0.3)$ $\quad\quad$ $\Delta_2 = (\ 0.5, 0.07)$

d) $\Delta_1 = (-0.2, 0.3)$ $\quad\quad$ $\Delta_2 = (\ 0.5, 0.3)$

Figures 4, 5 and 6 illustrate the robustness of B_1-, B_2- and B_3-optimal designs based on numerical optimization for these four pairs of priors. Note that in all cases the new criteria yield designs with less extreme efficiencies compared to those for the B-optimal design.

5. DISCUSSION

The criteria given in section 3 are applicable to any problem with any set of priors. Finding optimal designs for any of the criteria will typically require numerical methods, but with today's desktop workstations that is not a problem. The results for the turning point problem illustrate that B_1-, B_2- and B_3-optimal designs are compromises which maintain reasonably high overall efficiency across a range of priors. We are currently applying these criteria to other problems, specifically linear logistic regression and nonlinear regression.

REFERENCES

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag: New York.

Chaloner K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Communication in Statistics-Theory and Methods*, 18, 1385-1400.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21, 191-208.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24, 49-54.

Cook, R. D. and Nachtsheim, C. J. (1982). Model robust, linear-optimal designs. *Technometrics*, 24, 49-54.

Lau, T. (1988). D-optimal designs on the unit q-ball. *Journal of Statistical Planning and Inference*, 19, 299-315.

Mandal, N. K. (1978). On estimation of the maximal point of a single factor quadratic response function. *Calcutta Statistical Association Bulletin*, 27, 119-125.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308-313.

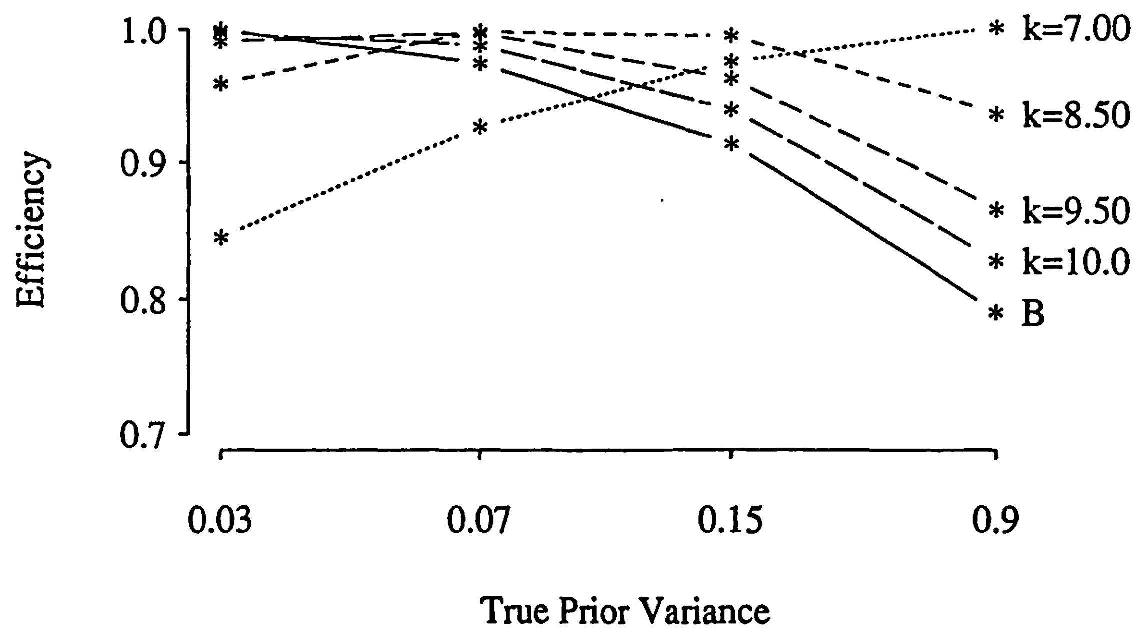
Silvey, S. D. (1980). *Optimal Design*, Chapman & Hall: London.

Stigler, S. M. (1971). Optimal experimental design for polynomial regression. *Journal of America Statistical Association*, 66, 311-318.

Studden, W. J. (1982). Some robust-type D-optimal designs in polynomial regression. *Journal of America Statistical Association*, 77, 916-921.

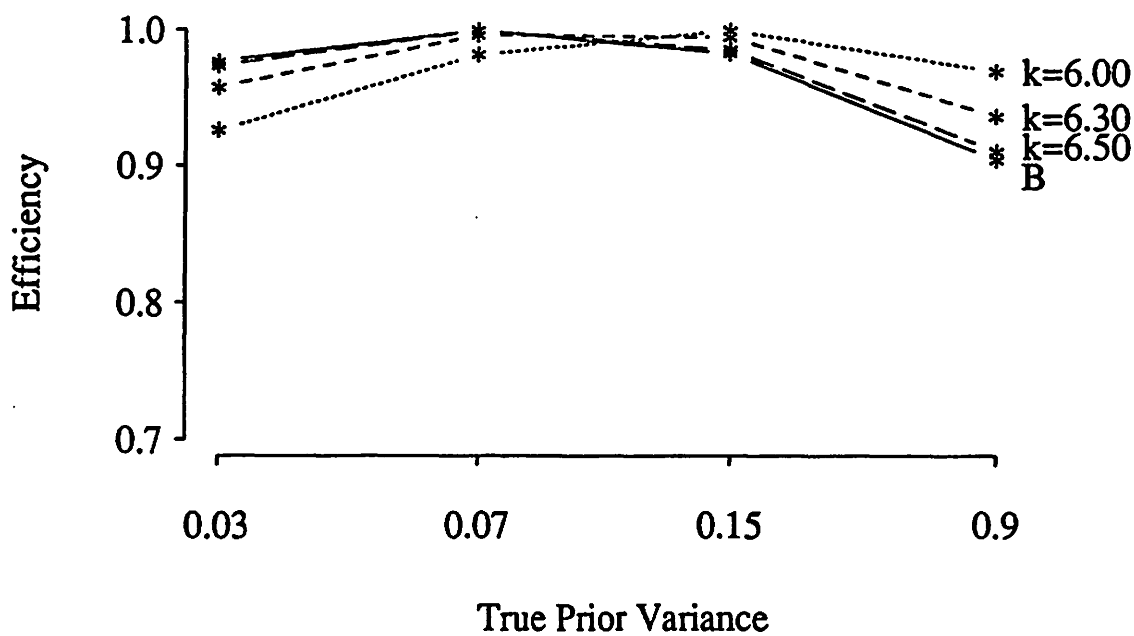
Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of Royal Statistical Society*, Series B, 35, 123-130.

(a)



Efficiency
1.0
0.9
0.8
0.7
0.03
0.07
0.15
0.9
True Prior Variance
k=7.00
k=8.50
k=9.50
k=10.0
B

(b)



Efficiency
1.0
0.9
0.8
0.7
0.03
0.07
0.15
0.9
True Prior Variance
k=6.00
k=6.30
k=6.50
B

(c)

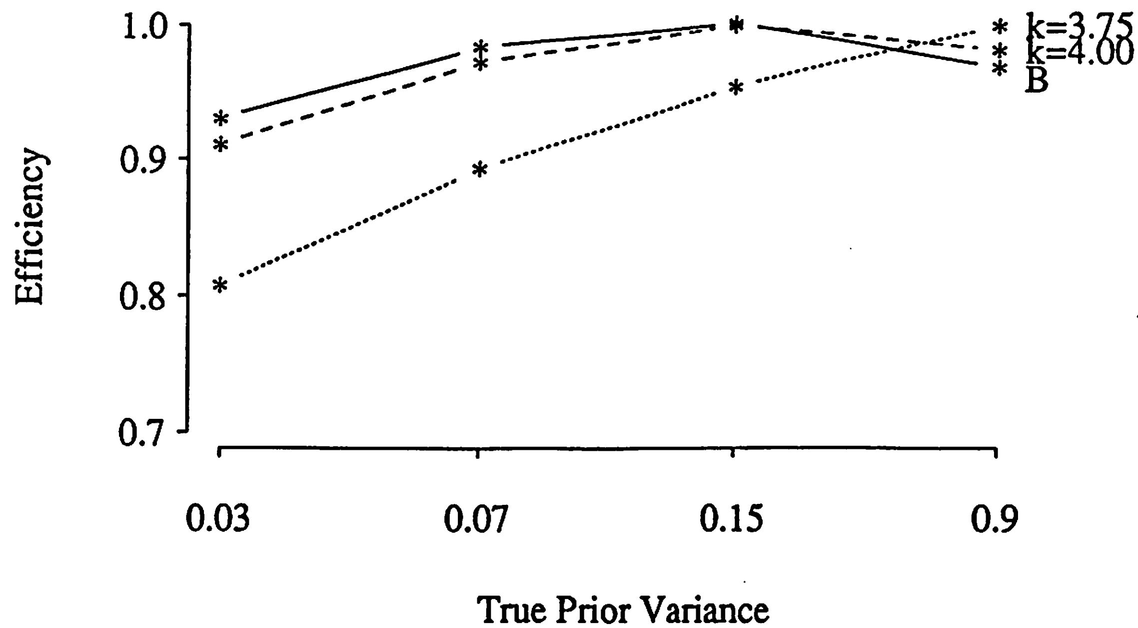
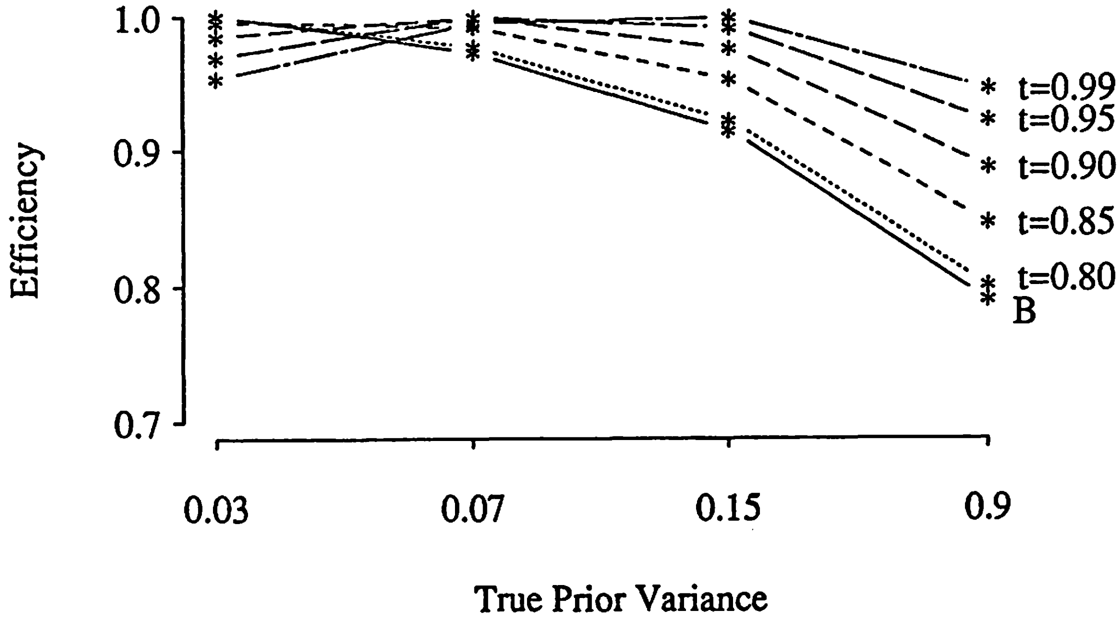
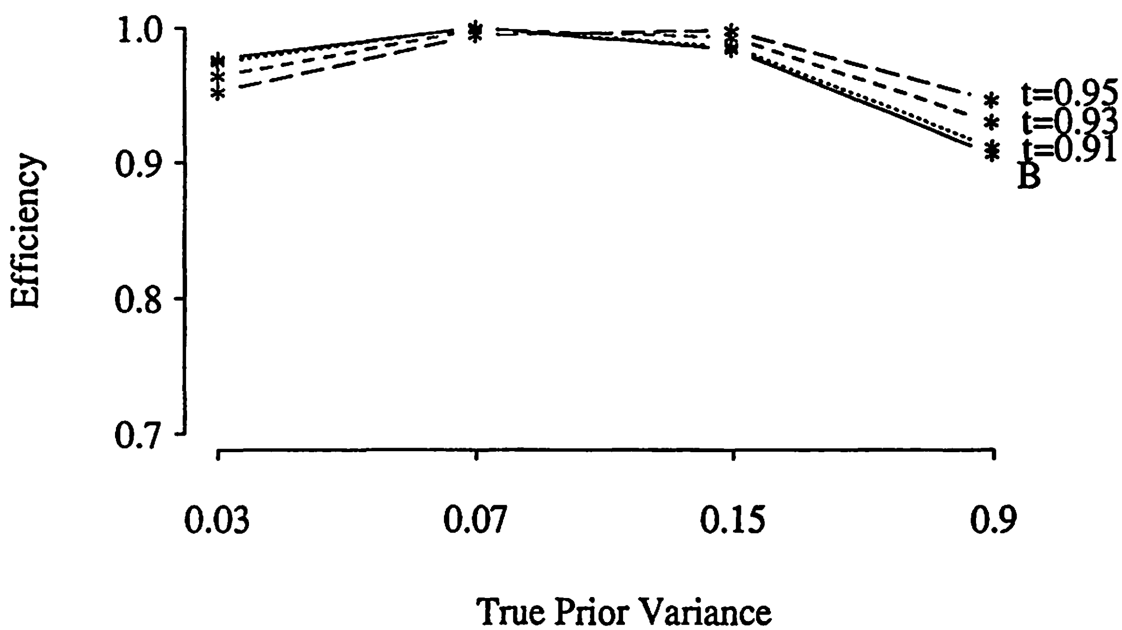


Fig 1. Efficiencies of B_1-optimal designs compared to B-optimal designs based on analytic results for priors Δ=(0, 0.03), Δ=(0, 0.07), Δ=(0, 0.15) and Δ=(0, 0.90) with major prior taken to be (a) Δ_1=(0, 0.03), (b) Δ_1=(0, 0.07) and (c) Δ_1=(0, 0.15).

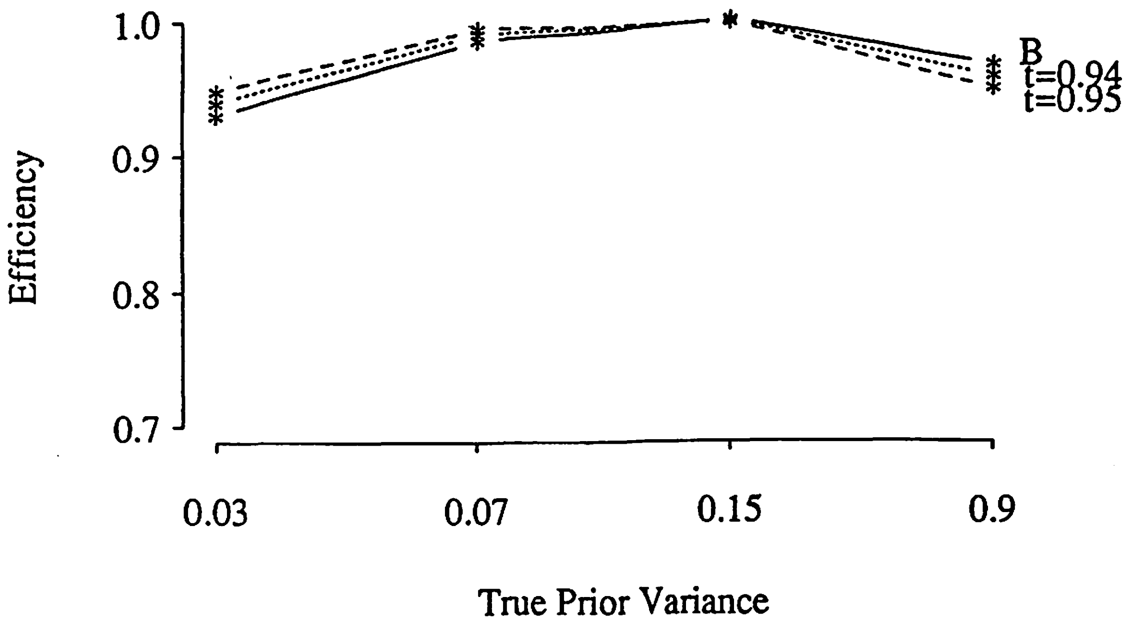
(a)


1.0
0.9
0.8
0.7
Efficiency
t=0.99
t=0.95
t=0.90
t=0.85
t=0.80
B
0.03
0.07
0.15
0.9
True Prior Variance

(b)


1.0
0.9
0.8
0.7
Efficiency
t=0.95
t=0.93
t=0.91
B
0.03
0.07
0.15
0.9
True Prior Variance

(c)



(d)

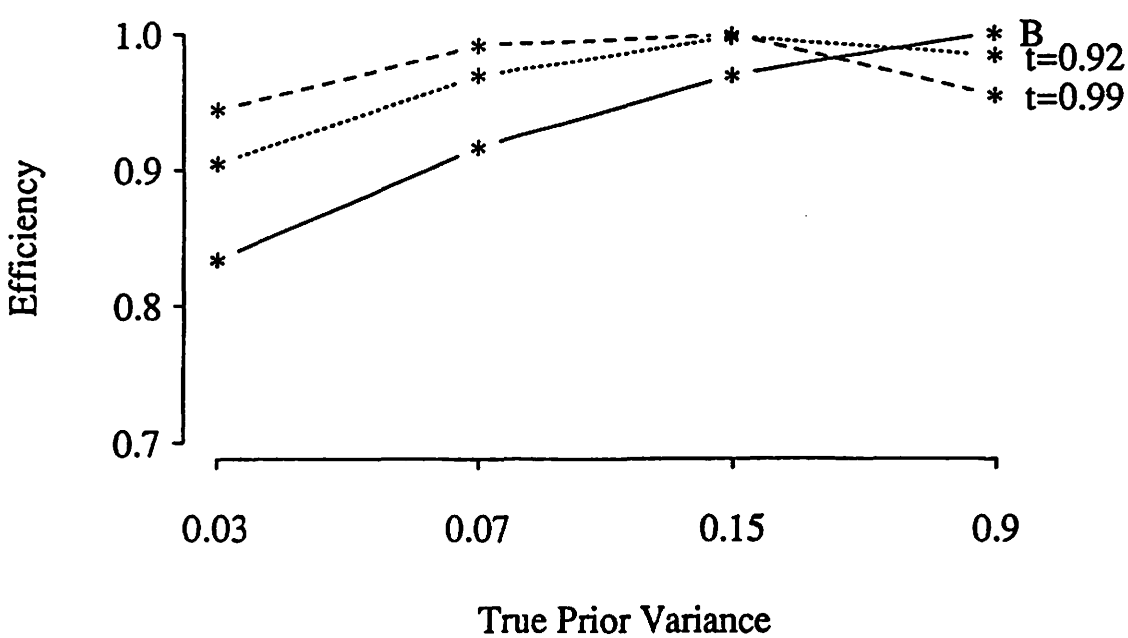
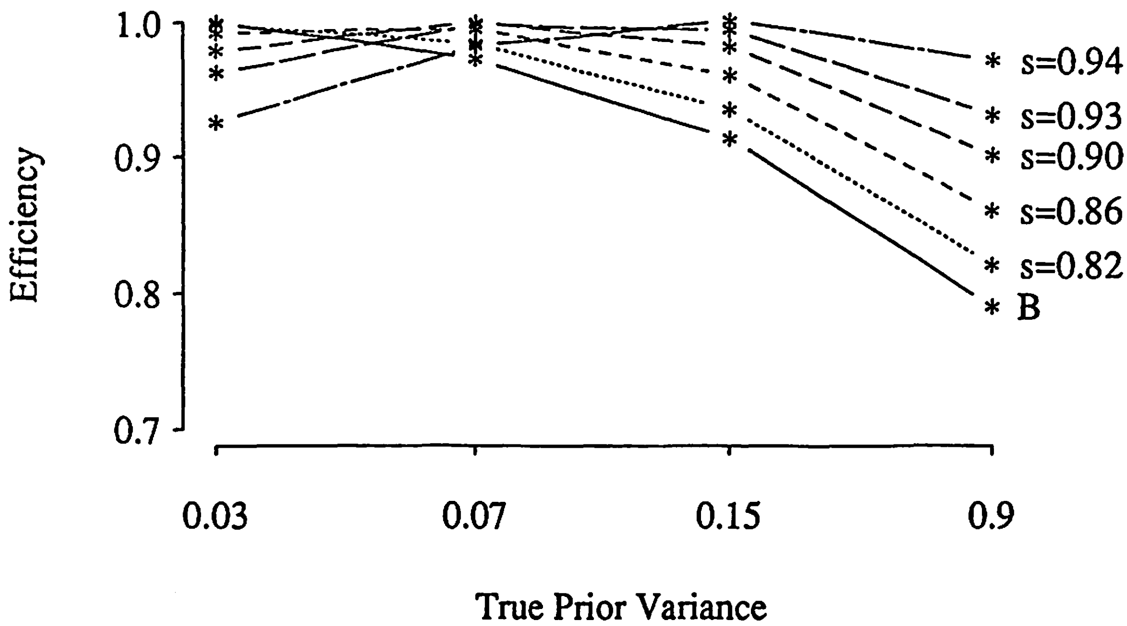
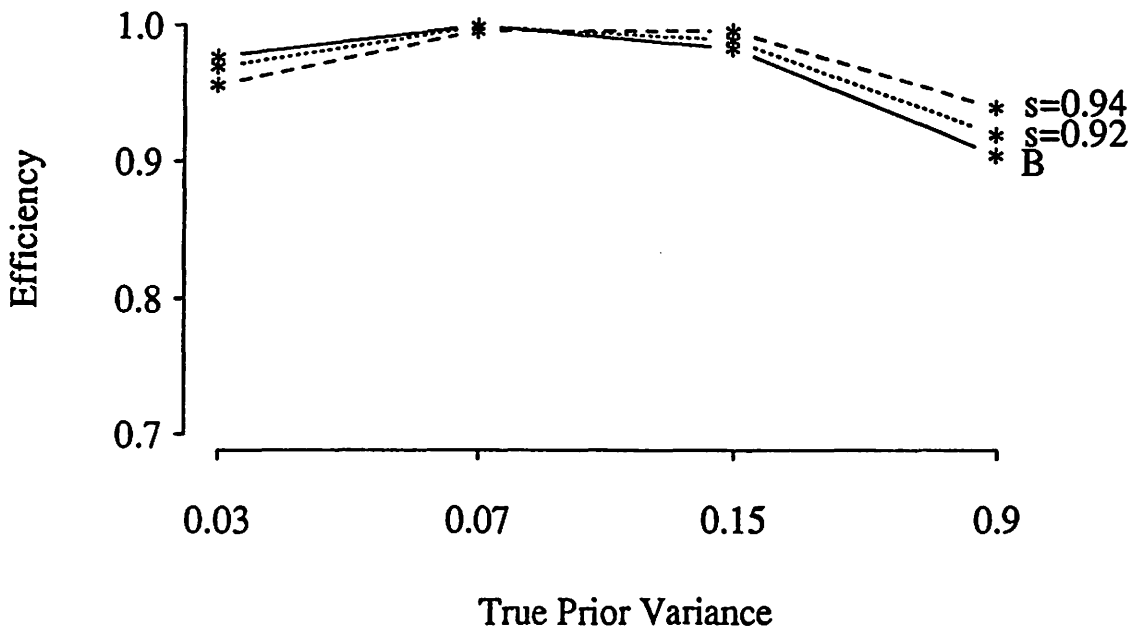


Fig 2. Efficiencies of B_2-optimal designs compared to B-optimal designs based on analytic results for priors $\Delta=(0, 0.03)$, $\Delta=(0, 0.07)$, $\Delta=(0, 0.15)$ and $\Delta=(0, 0.90)$ with major prior taken to be (a) $\Delta_1=(0, 0.03)$, (b) $\Delta_1=(0, 0.07)$, (c) $\Delta_1=(0, 0.15)$ and (d) $\Delta_1=(0, 0.90)$.

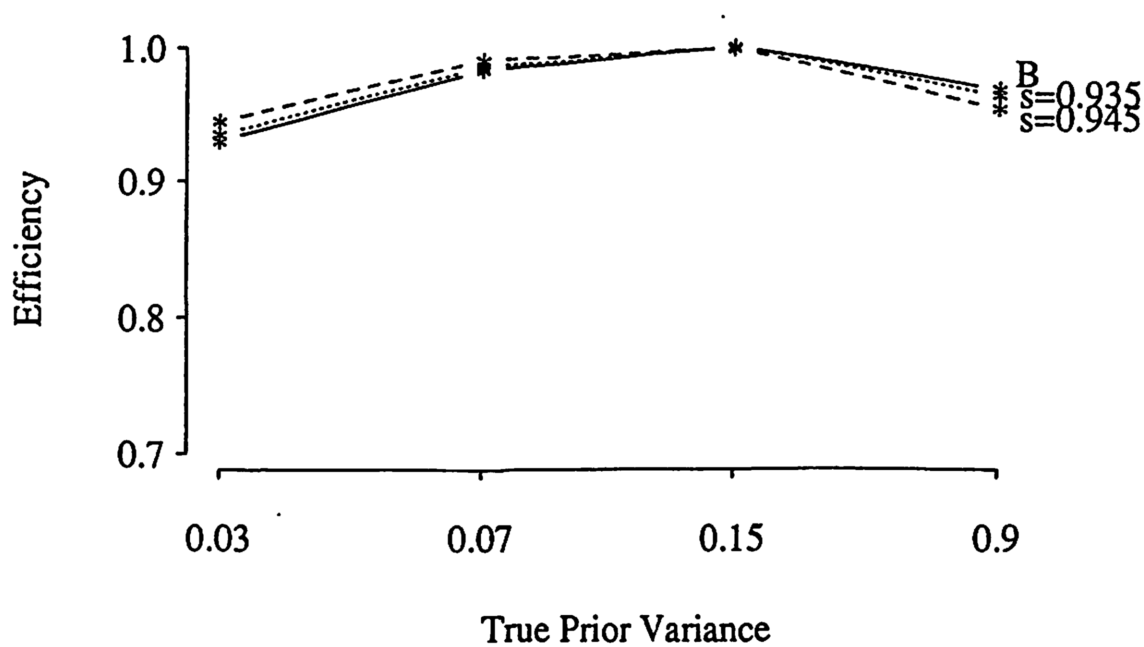
(a)


1.0
0.9
0.8
0.7
Efficiency
s=0.94
s=0.93
s=0.90
s=0.86
s=0.82
B
0.03
0.07
0.15
0.9
True Prior Variance

(b)


1.0
0.9
0.8
0.7
Efficiency
s=0.94
s=0.92
B
0.03
0.07
0.15
0.9
True Prior Variance

(c)



(d)

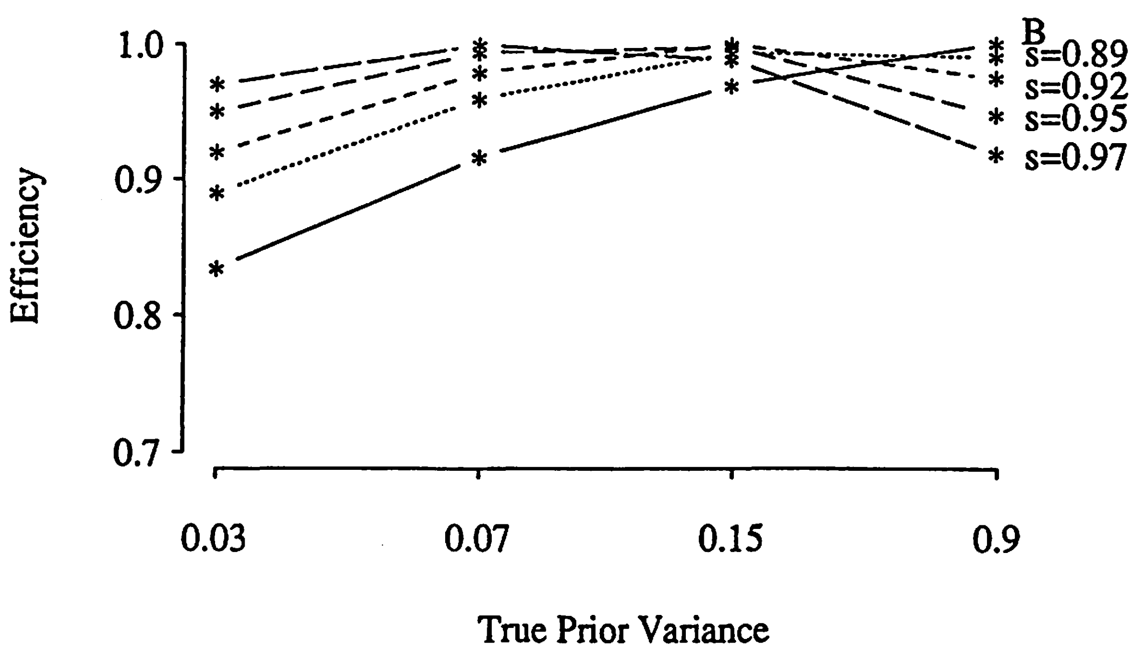
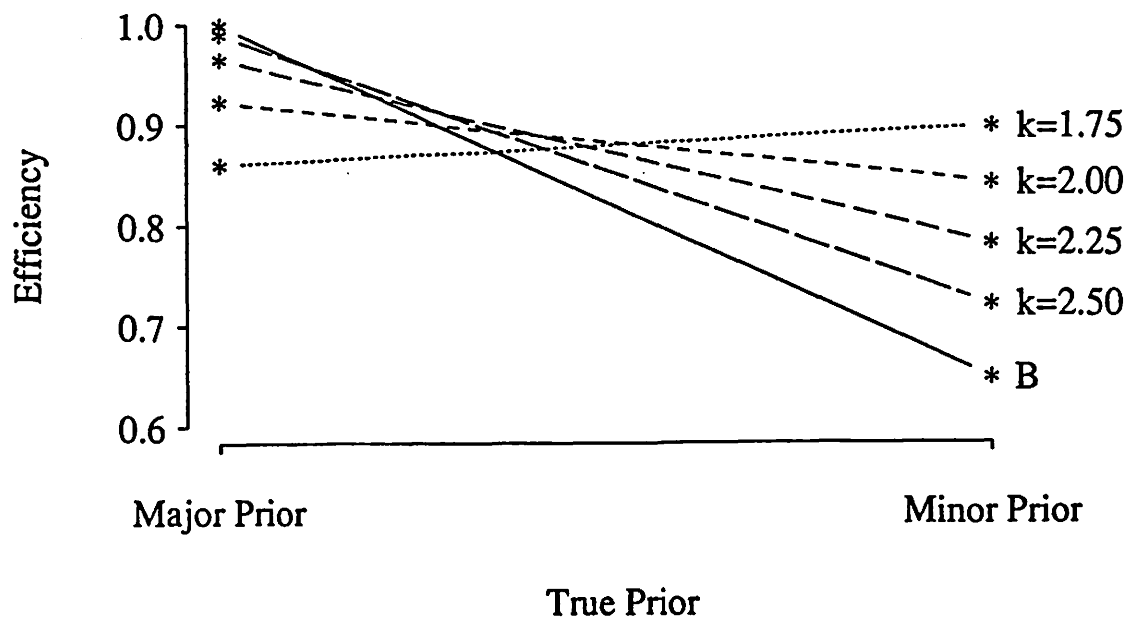
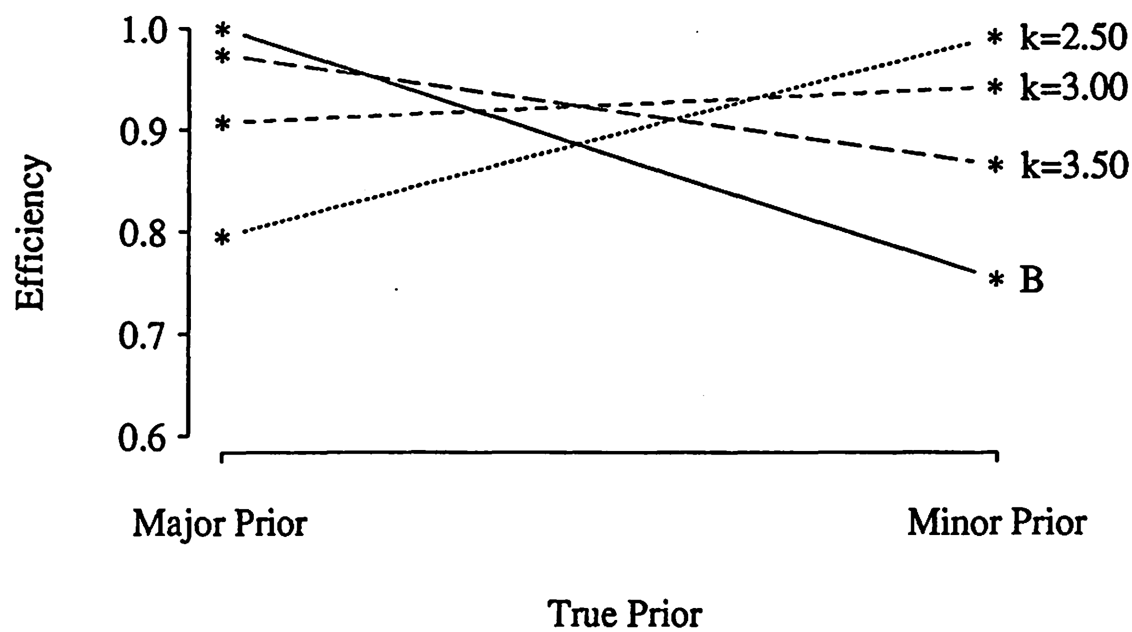


Fig 3. Efficiencies of B_3-optimal designs compared to B-optimal designs based on analytic results for priors $\Delta=(0, 0.03)$, $\Delta=(0, 0.07)$, $\Delta=(0, 0.15)$ and $\Delta=(0, 0.90)$ with major prior taken to be (a) $\Delta_1=(0, 0.03)$, (b) $\Delta_1=(0, 0.07)$, (c) $\Delta_1=(0, 0.15)$ and (d) $\Delta_1=(0, 0.90)$.

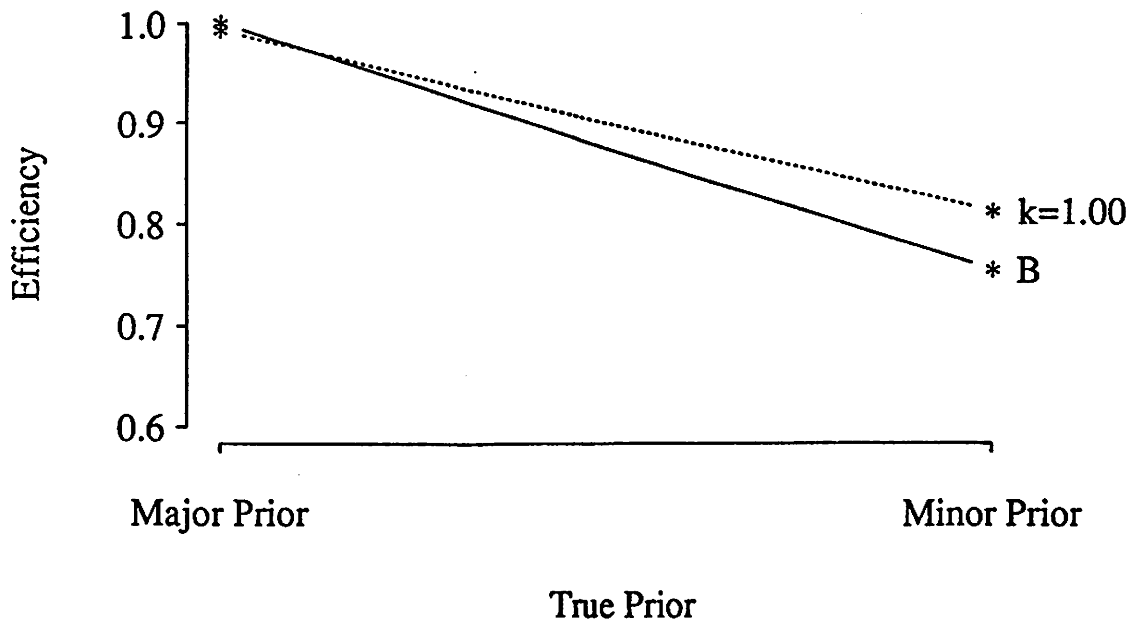
(a)


1.0
0.9
0.8
0.7
0.6
Efficiency
k=1.75
k=2.00
k=2.25
k=2.50
B
Major Prior
Minor Prior
True Prior

(b)


1.0
0.9
0.8
0.7
0.6
Efficiency
k=2.50
k=3.00
k=3.50
B
Major Prior
Minor Prior
True Prior

(c)



(d)

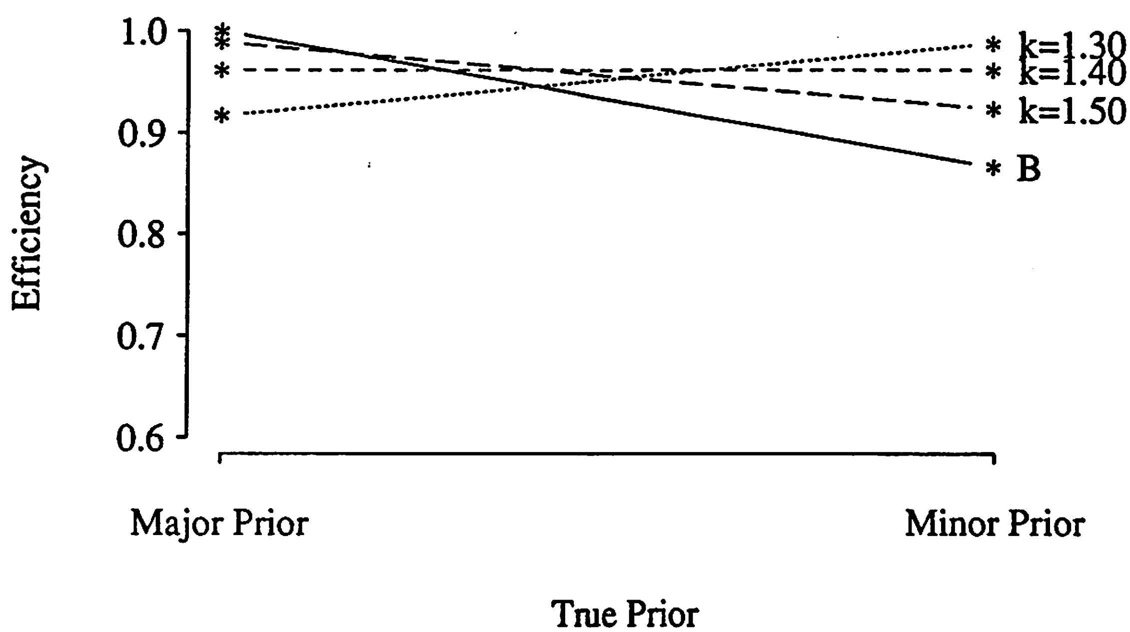
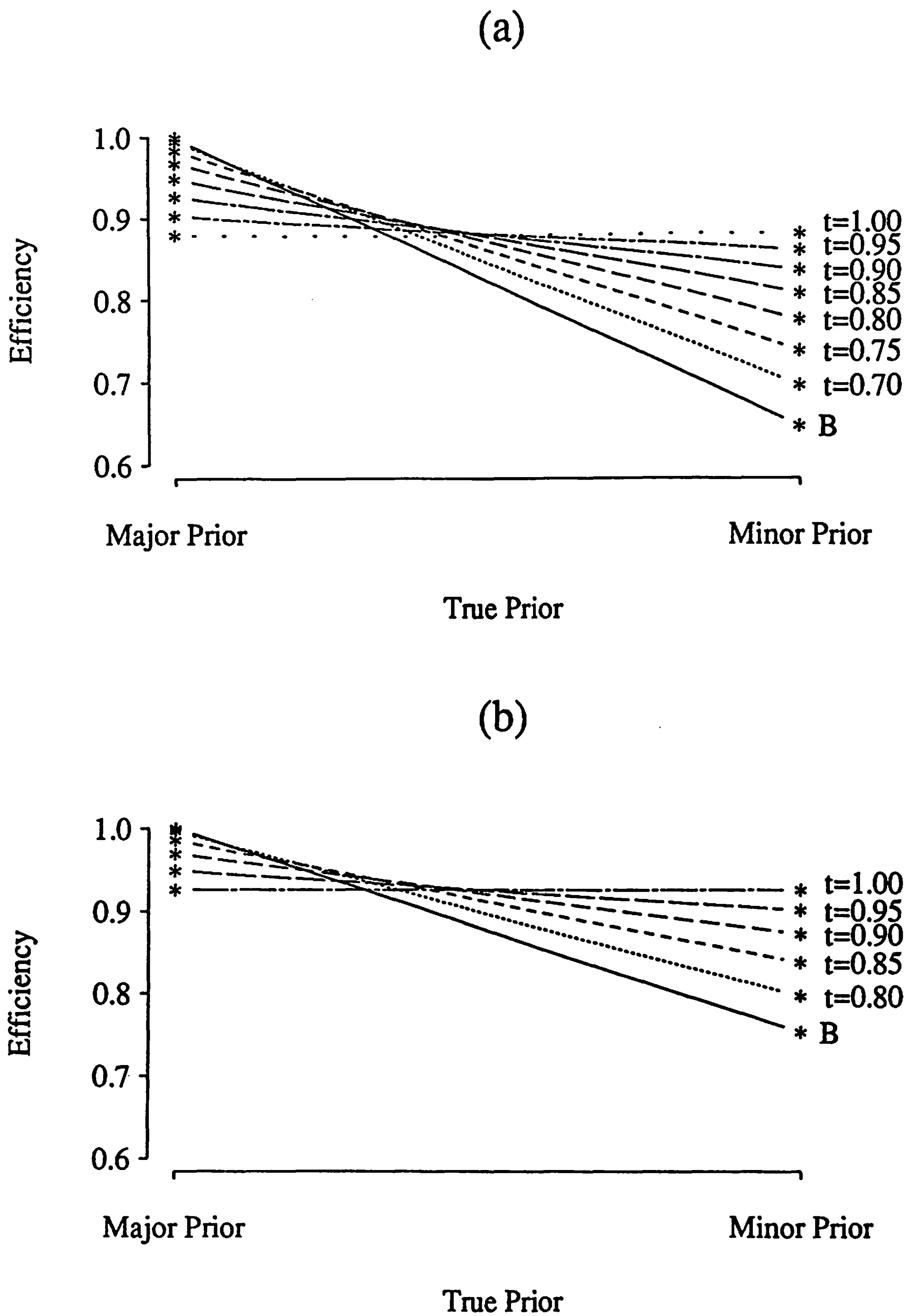
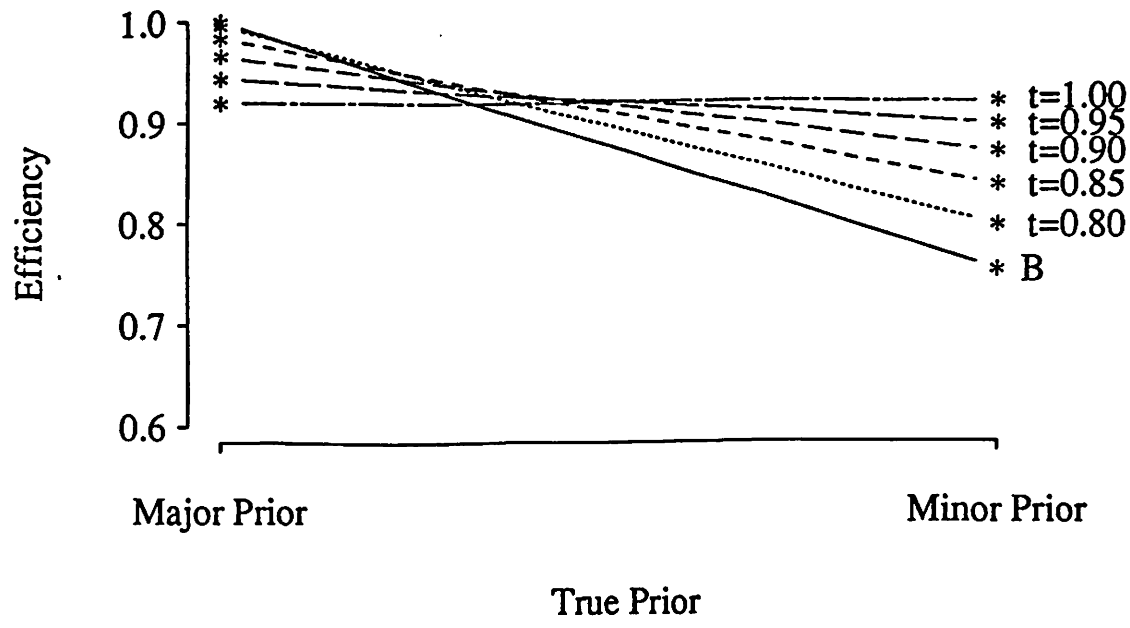


Fig 4. Efficiencies of B_1-optimal designs compared to B-optimal designs based on numerical results for designs with priors (a) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.07), (b) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.3), (c) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.07) and (d) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.3).


(a)
1.0
0.9
0.8
0.7
0.6
Efficiency
t=1.00
t=0.95
t=0.90
t=0.85
t=0.80
t=0.75
t=0.70
B
Major Prior
Minor Prior
True Prior
(b)
1.0
0.9
0.8
0.7
0.6
Efficiency
t=1.00
t=0.95
t=0.90
t=0.85
t=0.80
B
Major Prior
Minor Prior
True Prior

(c)



(d)

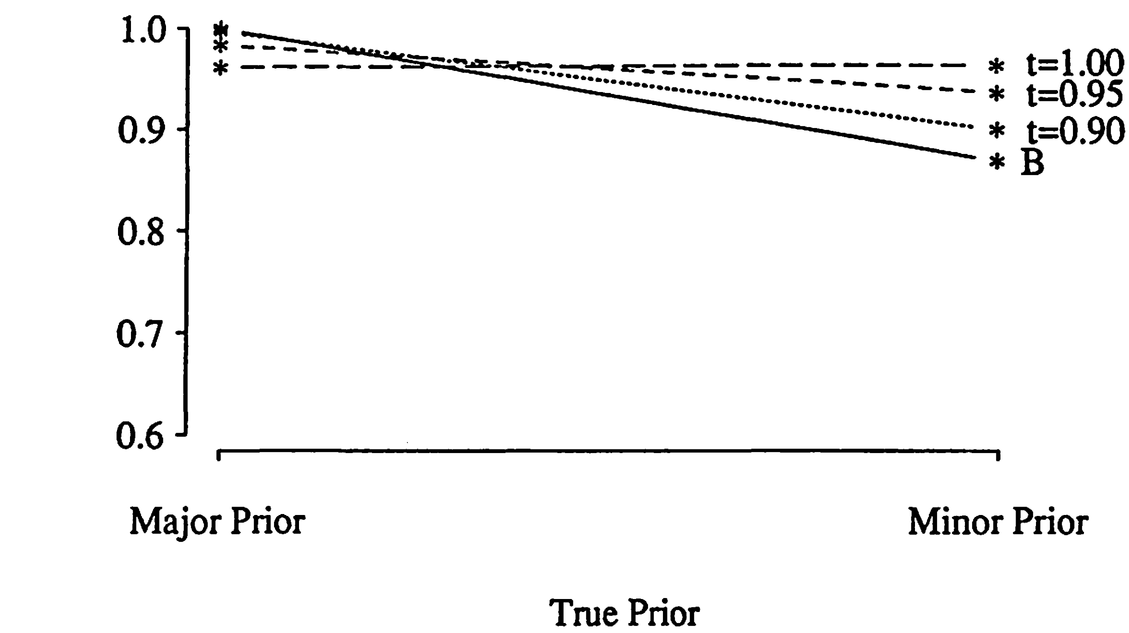
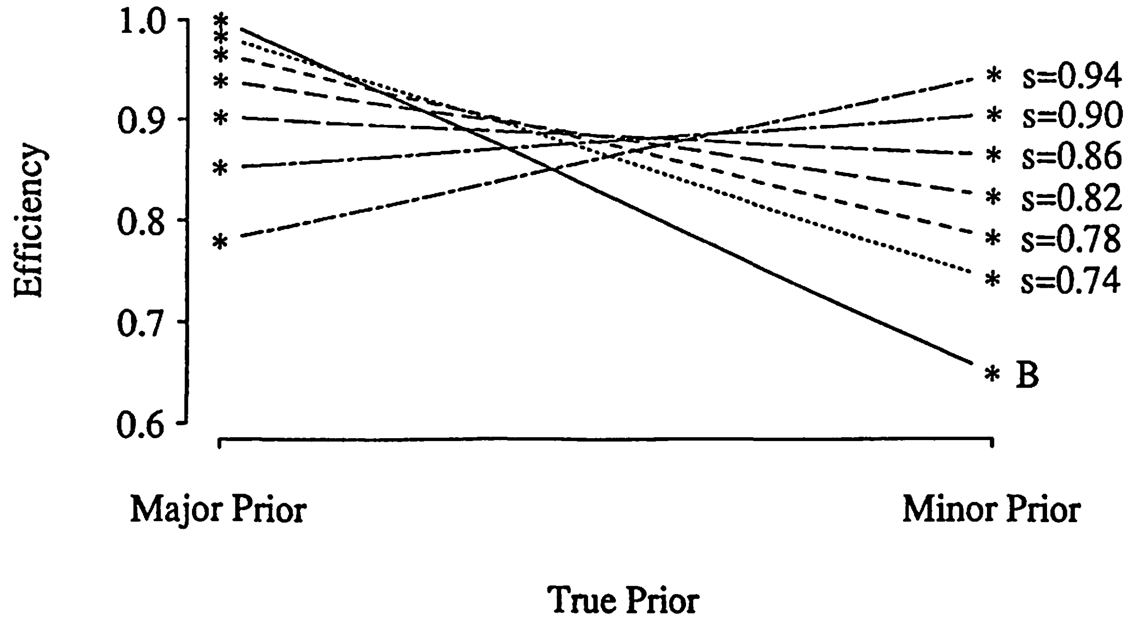


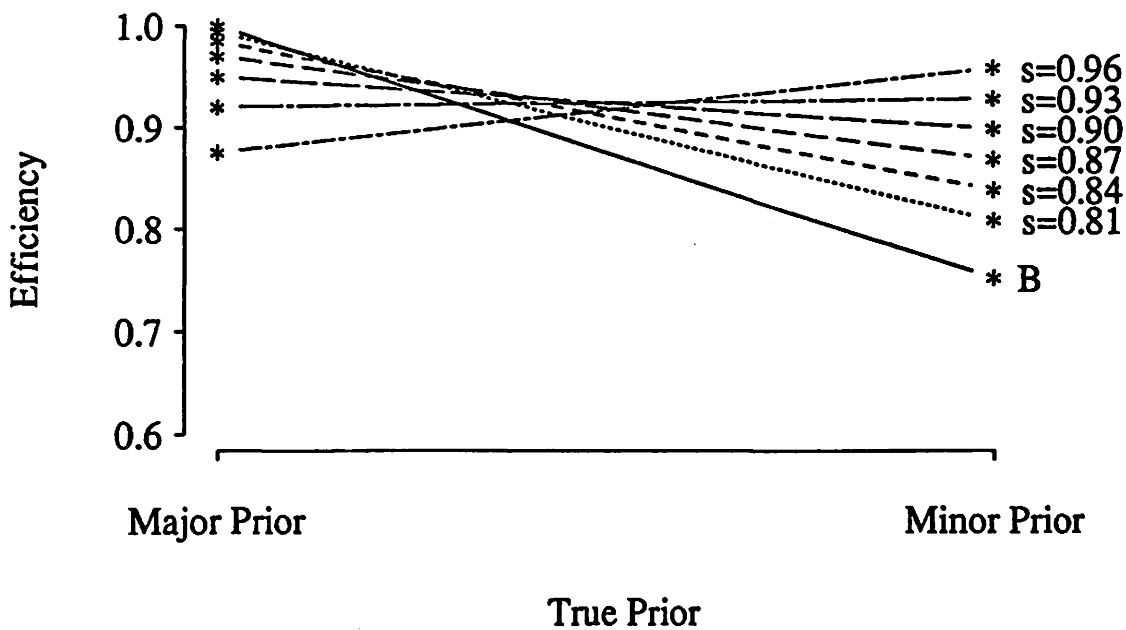
Fig 5. Efficiencies of B_2-optimal designs compared to B-optimal designs based on numerical results for designs with priors (a) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.07), (b) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.3), (c) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.07) and (d) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.3).

(a)



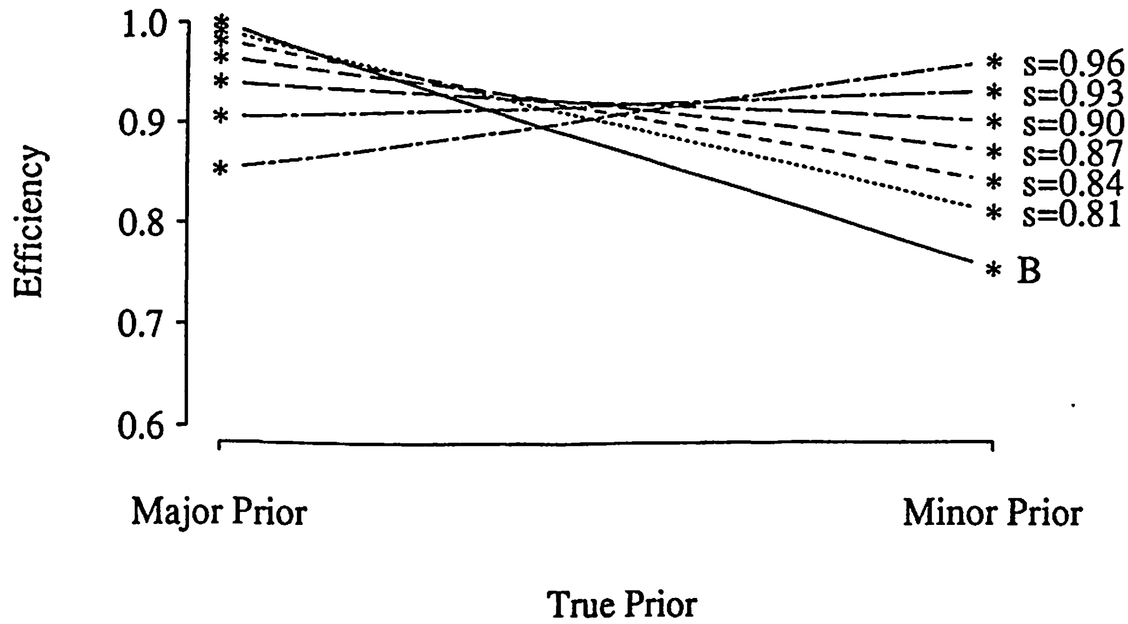
Efficiency
1.0
0.9
0.8
0.7
0.6
s=0.94
s=0.90
s=0.86
s=0.82
s=0.78
s=0.74
B
Major Prior
Minor Prior
True Prior

(b)



Efficiency
1.0
0.9
0.8
0.7
0.6
s=0.96
s=0.93
s=0.90
s=0.87
s=0.84
s=0.81
B
Major Prior
Minor Prior
True Prior

(c)



(d)

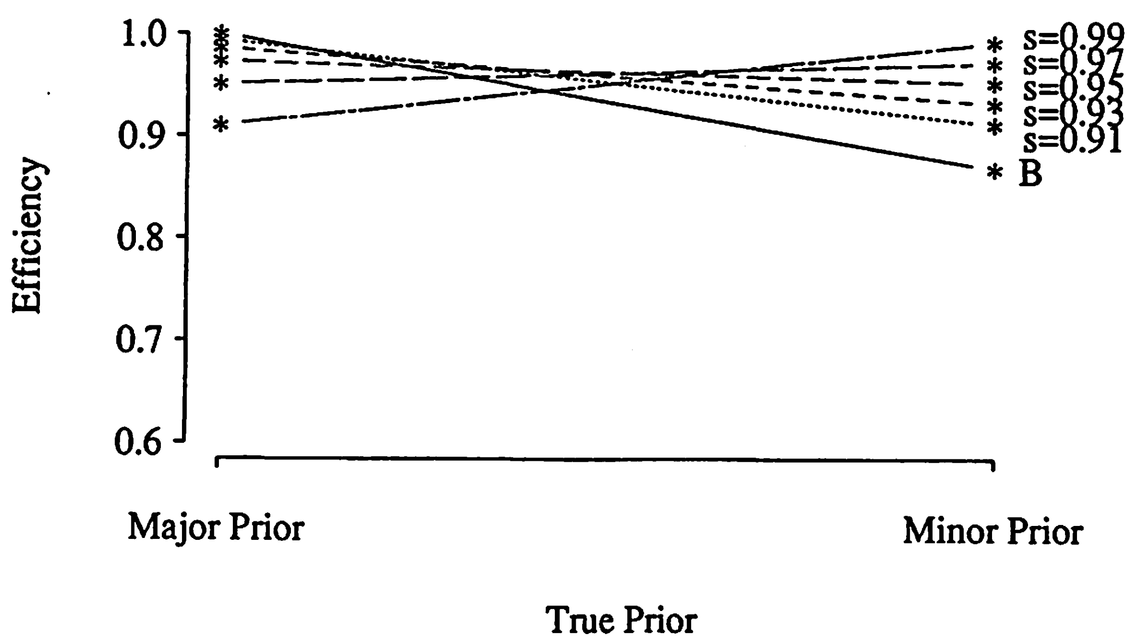


Fig 6. Efficiencies of B_3-optimal designs compared to B-optimal designs based on numerical results for designs with priors (a) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.07), (b) Δ_1=(-0.2, 0.07) Δ_2=(0.5, 0.3), (c) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.07) and (d) Δ_1=(-0.2, 0.3) Δ_2=(0.5, 0.3).